Modeling the Effects of Students' Interactions with Immersive Simulations using Markov Switching Systems

Nicholas Hoernle
Harvard University
nhoernle@g.harvard.edu

Kobi Gal
Ben-Gurion University
kobig@bgu.ac.il

Barbara Grosz
Harvard University
barbara@eecs.harvard.edu

Pavlos Protopapas
Harvard University
pavlos@seas.harvard.edu

Andee Rubin
TERC
andee_rubin@terc.edu

ABSTRACT

Simulations that combine real world components with interactive digital media provide a rich setting for students with the potential to assist knowledge building and understanding of complex physical processes. This paper addresses the problem of modeling the effects of multiple students' simultaneous interactions on the complex and exploratory environments such simulations provide. We work towards assisting educators with the difficult task of interpreting student exploration. We represent the system dynamics that result from student actions with a complex time series and use switch based models to decompose the time series into individual periods that target interpretability for teachers. The model learns the transition points between successive periods in the time series as well as the internal dynamics that govern each period. This model differs from other switch based models in that it decomposes the time series in a way that is human interpretable. This approach was applied to data that was obtained from an existing multi-person simulation with pedagogical goals of teaching sustainability and systems thinking. A visualization of the model was designed to validate the interpretability of the generated text-based descriptions when compared to a movie representation of the system dynamics. A pilot study using this visualization indicates that the switch based model finds relevant boundaries between salient periods of student work.

Keywords

Bayesian Inference, Exploratory Learning Environment, Markov Chain Monte Carlo, Interpretability, Switching State Space Models

1. INTRODUCTION

Complex systems simulations are becoming increasingly common in formal and informal STEM learning environments [21]. These simulations present scientific phenomena in a manner that bridges principles of science and the firsthand experience of emergent, real-world outcomes. However, the open-ended and exploratory nature of these simulations presents challenges to teachers' understanding of students' learning. Students' actions have immediate and long-term effects on the simulation leading to a rich array of emergent outcomes. Teachers may wish to discuss students' interactions to highlight salient learning opportunities, but if there are too many "moving parts" to the simulation, this becomes a challenging ideal.

This paper describes an automatic method for extracting salient periods from the log files that are generated by complex exploratory learning environments (ELE). Our goal is to generate relevant summaries of the system dynamics such that teachers can effectively engage students in discussions that stem from their own experiences with the simulations. We study an application of Switching State Space Models (SSSM) to the task of extracting salient periods from a mixed reality ELE, Connected Worlds, installed at the New York Hall of Science (NYSci). SSSMs [7] are a class of model for time-series data where the parameters controlling a linear dynamic system switch according to a discrete latent process. These models have seen use in a wide variety of domains including control [11], statistics [2], econometrics [8] and signal processing [14]. SSSMs combine hidden Markov and state space models to capture *regime* switching in non-linear, continuous valued time series [22]. The intuition is that a system evolves over time but may undergo a regime change that results in an intrinsic shift in the system's characteristics. Allowing for discrete points in time where the dynamics change, enhances the power of the simple linear models to capture more complicated dynamics. We propose that regime switching models also help to increase the *interpretability* of large and complex systems by automatically segmenting a time series into regions of approximately uniform dynamics. The result is that a complex session is broken into smaller periods that are more readily understood upon reflection on the session.

In this paper we introduce the Connected Worlds ELE and explain why teachers might need assistance when leading a discussion with the students where they reflect upon their actions. We expound on the SSSM and propose a method for decomposing a complex time series into smaller periods aiming to assist teachers when reflecting on a session with a class. We lastly present results showing the efficacy of

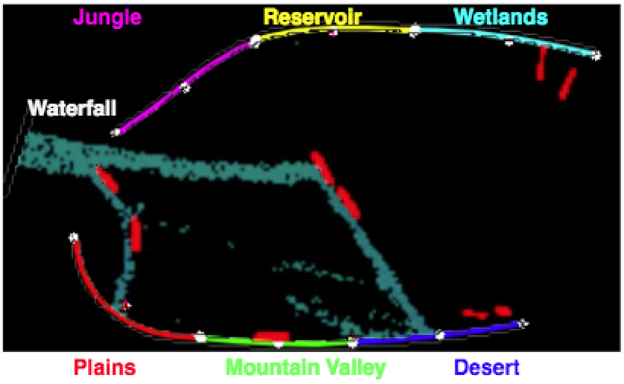

Figure 1: Bird's eye snapshot view taken from the movie representation of the CW environment. Biomes are labeled on the perimeter and logs appear as thick red lines. Water enters via the waterfall and in this image it mainly flows toward the desert and the plains.

our approach on both synthetic data and on data collected from CW. The CW validation is a preliminary study with significant results which suggest that the model output is human interpretable.

2. CONNECTED WORLDS

Connected Worlds[1] (CW) is a multi-person ecology simulation with the goal of teaching students about complex systems and systems thinking. It consists of an immersive environment comprising four interconnected biomes connected by a central flow of water that is fed by a waterfall. The simulation exhibits large scale feedback loops and presents the opportunity for participants to experience how their actions can have (often unintended) effects that are significantly removed in time and-or space. Students plant trees which flourish or die, animals arrive or depart, and rain clouds form, move through the sky and deposit rain into the waterfall.

Students interact with CW by positioning logs to control the direction of the water that flows in the simulation. Water can be directed to each of the four biomes (desert, plains, jungle, wetlands) and the distribution of flowing water depends on the placement of the logs. Water enters the simulation in two ways. The students can actively release water into the system from the stored water in the reservoir. Rainfall events are out of the students' control and these release water into the waterfall (to replenish the primary source of water) and into the individual biomes.

Figure 1 shows a bird's eye snapshot view of the state of the simulation in CW. The nature of the simulation is complex on a variety of dimensions. The simulation involves a large number of students simultaneously executing actions that change the state of the simulated environment. No one person - including the teacher or interpreter - can possibly follow what happens, even in a relatively short simulation. Each participant will have a different view of what transpired, depending on the actions s/he took and the state changes that resulted. Thus it is important to develop tools that can support teachers' understanding of the effects of students' interactions in complex ELEs such as CW.

3. RELATED WORK

This work is related to two separate strands of research: studying students' interactions in mixed reality ELEs, and modeling complex systems using switching models.

There is increasing evidence of the value of multi-person participatory simulations for engaging learners with complex science topics [9, 1, 23]. Research has explored classroom-scale participatory simulations where students play active roles in the simulation. Some examples include topics in disease transmission [3] and human body systems [12]. Other work has placed students in the role of scientists experimenting with simulated ecosystems [17, 4]. Within all of these examples, learners both engaged directly with the simulation during enactment, and reflected on their actions afterward to better understand how their choices resulted in observed system outcomes. Research has shown that using data obtained from students' own performances has the potential to engage them more effectively than presenting them with the results of an abstract simulation [16, 15]. Building on this work, our eventual goal is to provide assistive tools for teachers to further enhance the pedagogical impact that such ELEs can achieve.

Much work has been completed in the field of mining meaningful knowledge from time series data [5, 10, 19]. Ghahramani and Hinton [7] introduce and give a detailed presentation of the SSSM. We adapt this model to the special structure that is inherent in CW. Cappé et al. [2] and Giordani et al. [8] use switching models to capture non-linear behavior in a time series. SSSMs have been effectively applied in object tracking domains where it is necessary to predict the trajectory of various objects. Whiteley et al. [22] introduce a sequential Monte Carlo algorithm for inference over switching state space models using discrete particle filters. We present a new avenue of study in which SSSM models are used to describe complex time series in a way that can be easily interpreted by people.

4. SWITCHING STATE SPACE MODELS

SSSMs are commonly used to describe time series[2] with non-linear dynamics in econometrics and signal processing applications [8, 14]. A SSSM includes M latent continuous valued state space models and a discrete valued switching variable. Each of the models, which we refer to as regimes, have their own dynamics. At each point in time, the switching variable selects one of the individual state-space models to generate an observation vector.

The SSSM is formalized as:

$$\begin{aligned} \mathbf{X_t^{(m)}} &= \Phi^{(m)}\mathbf{X_{t-1}^{(m)}} + w_t^{(m)} \\ \mathbf{Y_t} &= S_t A^{(m)} \mathbf{X_t^{(m)}} + v_t \end{aligned} \quad (1)$$

Here, $X_t^{(m)}$ denotes the latent continuous valued state for

[1]https://nysci.org/home/exhibits/connected-worlds/

[2]Refer to Shumway and Stoffer [20] for a detailed discussion of time series analysis models.

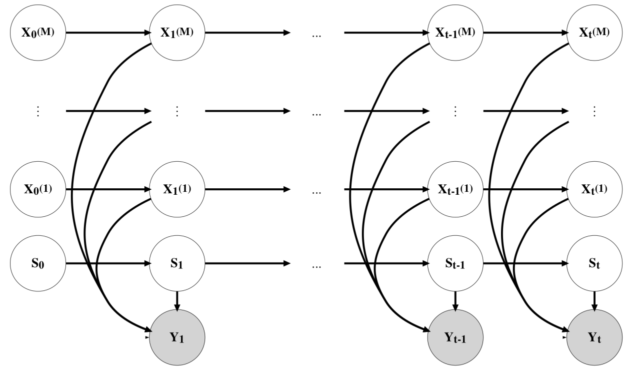

Figure 2: Graphical model for the switching-state space model. A latent discrete switching variable (S_t) selects an active, continuous state space model ($X_t^{(m)}$). The observation vector (Y_t) depends on the active regime at time t.

regime m at time t. S_t is a switching variable that selects the m^{th} regime such that regime m at time t produces observation vector Y_t, which depends on the latent state $X_t^{(m)}$. The states $X_t^{(m)}$ evolve over time in a way that depends on the transition matrix $\Phi^{(m)}$ and the previous state X_{t-1}. Figure 2 presents a graphical representation of an SSSM. Edges between variables represent stochastic casual relationships. Not shown in the figure are the regime dependent transition noise $w_t^{(m)}$ and the observation noise v_t. $A^{(m)}$ is the output matrix in the state space formulation, set to identity matrix I in our case.

We illustrate how an SSSM can describe the effects of students' interactions in CW. Y_t represents the observed water level in the different areas of the simulation at time t. $X_t^{(m)}$ describes the expected levels of water under regime m at time t. $\Phi^{(m)}$ controls the water flow in the simulation according to the transitions in regime m. S_t selects which of the regimes to use to describe the water level Y_t.

Importantly, a single regime is insufficient for modeling the effects of students' interactions with CW. This is because students' actions have a complex impact on the system dynamics. We therefore need to define multiple regimes, where each regime describes a series of events that can be (stochastically) explained by the regime dynamics. A regime is active for a duration of time in CW; we call this duration a period. For example, in one period water is mainly flowing to the plains and to the desert (as is shown in figure 1). In the next period, students move the logs to re-route water flow to the wetlands potentially because plant life is dying. Each of these periods might be active for different durations and their dynamics are described by different regimes.

4.1 Exploiting Model Structure

We aim to perform inference over the latent states, $X_t^{(m)}$, the regime parameters, $\Phi^{(m)}$, and latent switching variable, S_t. Computing posterior distributions for SSSM is computationally intractable [18]. To illustrate, in figure 2 we see that the graph consists of M state space models that are marginally independent. These models become conditionally dependent when Y_t is observed, as is the case in this graph. The result is that $X_t^{(m)}$ is conditionally dependent on the value of all of the other latent states and switching variables for times 1 through T and regimes 1 through M [18]. Previous approaches use approximate methods such as variational inference [7] and a 'merging of Gaussians' [14, 18] to address the inference problem. The variational inference approximation transforms the intractable Bayesian expectation problem into an optimization problem by minimizing the Kullback Leibler (KL) divergence between a simpler family of approximating distributions and the unknown, intractable posterior. The merging of Gaussians approach uses a single Gaussian to represent the mixture of M Gaussians at each time step thereby simplifying the computation with the cost of being susceptible to local optima (see section 5.1).

While these methods have seen success in previous examples, they cannot be applied to our domain. This is because they allow the system to switch back and forth between regimes, resulting in frequent regime changes that can hinder the interpretability of the model output. This work takes a different approach by imposing structure on the model to address both inference and interpretability challenges. Further, as the optimization procedures of the previous work are susceptible to local optima, we rather use a Markov chain Monte Carlo (MCMC) approach to approximate the posterior distribution of the latent parameters.

We make two assumptions, which arise from the need to create human interpretable descriptions of complex system behavior. **Assumption 1:** the system advances through a series of regimes, each regime is active for a period, and then switches to an entirely new regime, one that has not been used before. **Assumption 2:** the regime remains active for the maximum possible time for which it can be used to describe the period.

To illustrate, without making these assumptions there are M possible assignments of regimes for each time step, making a total of M^T combinations of possible assignments, which is exponential in the number of time steps. Moreover, in the worst case, the number of possible periods is bounded by T with a switch at every time step. In contrast, under our assumptions, there are only two possible assignments of regimes for each time step (i.e., do we stay in the current regime or do we progress to the next regime), making for a total of 2^M combinations of possible assignments, where M is constant. The number of possible periods under this methodology is bounded by M. We hypothesize that the forced reduction in complexity of the fitted model would significantly simplify the interpretability of the model for a human.

4.2 Algorithm for Posterior Inference

Computing the posteriors in an SSSM corresponds to approximating the joint distributions over $X_t^{(m)}$ and $\Phi^{(m)}$ given the observation vector $\mathbf{Y}$. A well known problem with MCMC inference in complex graphical models with hidden variables is that of identifiability [13]. Models are nonidentifiable when two sets of parameters can explain the observed data equally well. For example, in a simple Gaussian mixture model with means μ_0, μ_1 and covariances Σ_0, Σ_1, the marginal posterior distributions of the parameters are identical. A possible solution to the identifiability problem is to add constraints (e.g. enforcing $\mu_0 > \mu_1$). However, defining constraints in higher-dimensional domains is non-trivial.

Another solution for solving the identifiability problem is to provide labels for part of the data. This is termed semi-supervised learning and we incorporate this solution into our model. In the context of the CW domain, we can label observations as belonging to one regime or another. Let $S_{t,t+1,\ldots,t-1+K,t+K}$ be a consecutive set of K state variables such that S_t and S_{t+K} have known value assignments (regime m and regime $m+1$ respectively). The values for the state variables $S_{t+1,\ldots,t-1+K}$ are unknown. By Assumption 1, the switch between regimes m and $m+1$ occurs at some S_l where $t < l \leq t+K$. Therefore, the value of S_l determines the values for all of the unknown states as S_t is assigned to regime m for $t < l$ and it is assigned to regime $m+1$ for $t \geq l$.

We provide a sketch of this process in Algorithm 1. Step 1 initializes the M supervised switch variables, one per regime. The labeled switch variables are spaced uniformly in time and are assigned to regimes in increasing order according to Assumption 1. This uniform method for initialization can be justified by Assumption 2, in that any set of regimes that provides an interpretable model is sufficient. The number of expected time steps in each period is $K = T/M$, and there are $K-2$ unlabeled switch variables between each pair of switch variables assigned to regimes.

Step 2 performs MCMC sampling to approximate the posterior of the model[3]. For the case when the value of the switch variable is known, the posterior of $X_t^{(m)}$ can be directly sampled by following the structure of a state space model. In the case where the switch variable is unknown, we have a marginalization problem over the two possible values of S_t. For the hidden Markov model (HMM) structure this can be efficiently computed with the forward algorithm [20]. To formulate the HMM forward algorithm, we use the observation probabilities from the individual state space models in place of the emission probabilities of a standard HMM. Here, π_{S_i} refers to the belief of the state of the switching variable given the evidence up to that point in time.

Step 3 uses the regime specific parameters $\Phi^{(m)}$ to make a maximum likelihood assignment of an observation to a regime using the Viterbi algorithm [20], thereby specifying the value of $S_t \forall t \in [1:T]$.

Algorithm 1 is computed on an SSSM that implements Assumptions 1 and 2. Such a model is shown in figure 3. The model depicts a subset of the time series with K time steps from time t to time $t+K$. There are two supervised labels at the boundaries of the subset with the variable S_t assigned to regime m and variable S_{t+K} assigned to regime $m+1$. The unknown $K-2$ states in between are marginalized over such that the regime specific posteriors can still be approximated. This model is repeated for the $M-1$ switches in the data. The setup is flexible in that informative priors for the model noise and transition matrices can be specified (and

[3]Implemented using Stan MC (`http://mc-stan.org/`)

Algorithm 1: Posterior inference algorithm

Input: M (number of regimes), $\mathbf{Y}$ (vector of observations for T time steps).

1 Initialization: Label one datapoint per regime, leaving $T-(M+1)$ unlabeled datapoints.

2 MCMC Inference: Draw samples for $X_t^{(m)}, \Phi^{(m)}$ from the posterior distribution defined by the structured probability model:

for Y_t *in* $\mathbf{Y}$ **do**
- **if** $S_t = m$ *is known* **then**
  - sample from $P(X_t^{(m)}, \Phi^{(m)} \mid X_{t-1}, S_t = m, Y_t)$
- **else**
  - marginalize over S_t. Sample from $\sum_{i=m-1}^{m} \pi_{S_i} P(X_t^{(i)}, \Phi^{(i)} \mid X_{t-1}, S_t = i, Y_t)$

3 Posterior Inference: Use the posterior for regime parameters ($\Phi^{(m)}$) to run a Viterbi pass on the data $\mathbf{Y}$ to make a maximum likelihood assignment of the value of S_t to regime m (thereby learning the switching variables S_t).

Output: S_t (assignments to regimes), $\Phi^{(m)}$ (regime posterior distributions).

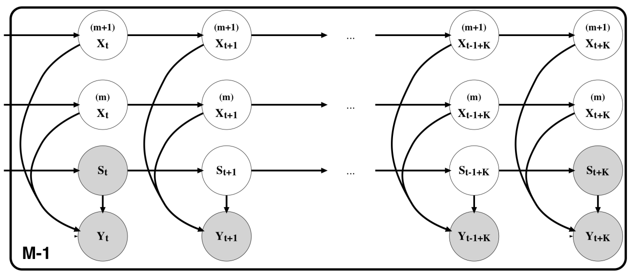

Figure 3: Updated graphical model showing the semi-supervised switching labels, along with the choice of only two chains between two semi-supervised points. This representation is repeated $M-1$ times to describe the $M-1$ switches between the M regimes.

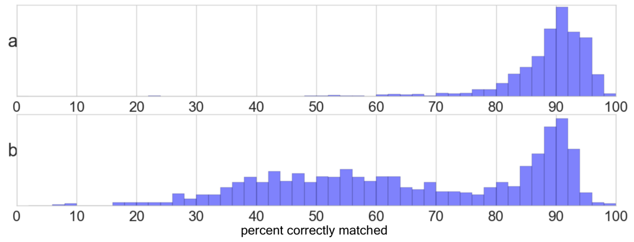

Figure 4: Histogram of the percent of correctly inferred labels for the observed output. The structured sampling Algorithm 1 (a) learns the regime labels more accurately than the randomly initialized Gaussian merging algorithm (b).

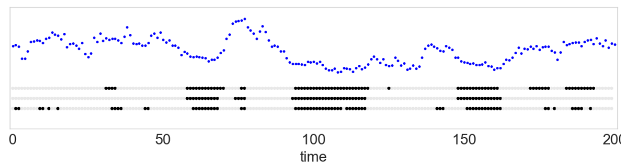

Figure 5: An example of a generated time series from the SSSM model of Equation 2. The x axis represents time, and the y axis shows the observations (the magnitudes of the signal are irrelevant for this investigation). Regime labels are shown as black and gray dots representing the two label options. True labels (top) are compared to the inferred labels from Algorithm 1 (middle) and the Gaussian merging (bottom).

related) as required by domain knowledge.

5. EMPIRICAL VALIDATION

We evaluate two aspects of Algorithm 1. First, we show that it finds the true regime labels in a synthetic dataset. Thereafter, we use data that were collected from Connected Worlds to run a preliminary experiment that tests whether the inferred periods are interpretable to human validators.

5.1 Evaluation on Synthetic Data

We generate synthetic data to test whether Algorithm 1 finds a reasonable representation of known switches in an SSSM. Equation 2 describes an SSSM with two regimes and a continuous state space. The transition parameters and regime noise are determined according to the active regime. This model is adapted from Ghahramani and Hinton [7] which describes a state space that is disjoint at regime switches; we rather chose to make the state space continuous at the switch points as this more accurately mimics the scenario that is present in CW. The prior probability of each of the regimes is 0.5 ($p_1 = p_2 = 0.5$); the regime transition probabilities are $S_{1,1} = S_{2,2} = 0.95$ and $S_{1,2} = S_{2,1} = 0.05$[4]. We used this model to generate 1000 time series, each with 200 observations.

$$\begin{aligned} \mathbf{X_t^{(1)}} &= 0.99\,\mathbf{X_{t-1}} + w_t^{(1)} \quad & w_t^{(1)} &\sim \mathcal{N}(0,1) \\ \mathbf{X_t^{(2)}} &= 0.9\,\mathbf{X_{t-1}} + w_t^{(2)} \quad & w_t^{(2)} &\sim \mathcal{N}(0,10) \\ \mathbf{Y_t} &= S_t\mathbf{X_t} + v_t \quad & v_t &\sim \mathcal{N}(0,0.1) \end{aligned} \tag{2}$$

We compare the Gaussian merging baseline that is commonly used in the literature [14] to Algorithm 1 with the number of regimes initialized to 9. The accuracy of each approach is measured as the percentage of the correctly labeled data points as belonging to either regime 1 or regime 2. On average Algorithm 1 labels 89% of the data correctly, materially higher than the 66% average accuracy of the Gaussian merging approach. Figure 4 shows a histogram of the correctly inferred switch points in the data according to Algorithm 1 (top) and the baseline (bottom). The bi-modal and long tailed distribution for the baseline approach demonstrates its susceptibility to local optima.

[4] $S_{j,k}$ denotes the probability of a switch from regime j to regime k.

Figure 5 shows an example of the generated time series (top) and the associated switch points (bottom). The switch points are shown according to the true model, the points inferred by Algorithm 1 and the points inferred by the baseline. Each period is represented by a sequence of black and gray colored circles. As shown by the figure, the periods inferred by Algorithm 1 and the baseline both overlap to some extent with the true periods. However, there is substantially more noise in the inferred periods of the baseline. Algorithm 1 learns the regime autoregressive parameters $\phi_1 = 0.97 \pm 0.027$ and $\phi_2 = 0.88 \pm 0.035$, again showing an effective recovery of the individual regime parameters.

The superior performance of Algorithm 1 can be directly attributed to the switching behavior that is enforced by Assumptions 1 and 2, which was not assumed by the baseline model. Although the model structure encourages the discovery of switches in Algorithm 1 the uniformly spaced labels should not be seen as a model advantage as no prior knowledge of the actual switches is used in performing this initialization step. Given that the proposed algorithm finds a reasonable representation for the switches in a generated dataset, we turn to the evaluation of the interpretability of its output within the CW context.

5.2 Preliminary Validation of Interpretability on Connected Worlds Data

Because the ultimate users of the output of Algorithm 1 will be teachers leading their students in a discussion of the simulation behavior, we wanted to confirm that the inferred switch points were interpretable by a human seeking to understand the “story” of the simulation. In order to do this, we used a movie of the water flow (see figure 1 for one such frame) and asked evaluators to select one of three possible switch points between every pair of consecutive periods. Evaluators saw a composite of 1) the movie of the two periods; 2) a description of the dynamics of each of the two periods and 3) a set of three possible switch points between the periods. The evaluator’s task was to choose the switch point that best matched the change in dynamics between the two periods. One of the three switch points was that inferred by Algorithm 1; the other two were random times sampled uniformly from the beginning of the first period to the end of the second period.[5]

The descriptions were generated from the inferred parameters that are an output from Algorithm 1. In equation 1, $\Phi^{(m)}$ refers to the transition matrix for the m^{th} regime. As is discussed in section 4, the parameters from this matrix describe the expected movement of water in the given period. We threshold the values from this matrix to generate a short text description for the water movement. One such description could be: *"Water is directed towards the desert and plains. The wetlands and jungle are receiving little or no water"*.

Evaluators worked with five sessions, each of which included 5 to 10 periods of system dynamics. Selecting the correct switch point is not a trivial task: it requires distinguishing between changes in the system that indicate different dynamic regimes and those that are noise within the same dynamic regime. We see an evaluator's ability to choose a switch point based on the movie and a description of the two contiguous periods as evidence that the inferred periods are usable by a teacher who wants to guide students in constructing a causal description of their experience with the simulation. Moreover, this can be seen as evidence that the inferred regime parameters match inferred period boundaries, together presenting a coherent description for the water movement for a short segment of the CW session.

Figure 6 shows the results of the validation using four evaluators with knowledge of the CW domain. The five sessions are shown along the x-axis; the fraction of correctly selected switch points is shown by the bin heights. The dashed line represents a random baseline in which the selected switch probability corresponds to $\frac{1}{3}$. Under the null hypothesis, the performance of an evaluator would not be significantly different than the random baseline. The results indicate that the evaluators chose the switch point identified by Algorithm 1 significantly more often than the random baseline ($p < 1 \times 10^{-4}$), suggesting that the inferred switch points were indeed interpretable to a large extent as meaningful changes in the state of the system. The differences in interpretability seen in figure 6 (e.g. session 4 was more difficult to interpret than session 3) can provide further guidance to us in how to support teachers and students in making sense of their experiences in CW. For example, the sessions with more complicated dynamics might need more periods to fully capture the progression over time. Predefining the number of periods for a given session is an aspect of this approach that needs addressing. A more detailed user study is left for future work.

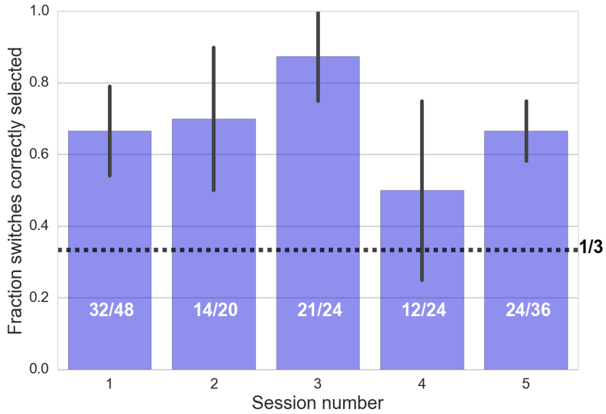

Figure 6: Expert validation of five different test files from sessions with CW. The histogram shows the fraction of correctly identified switches between automatically identified periods with an expected baseline accuracy of $\frac{1}{3}$.

6. CONCLUSION AND FUTURE WORK

This paper has presented novel research into the simplification of log files that are generated by complex participatory immersive simulations. The log files were represented as a time series that was decomposed with the long term goal of producing periods that are useful for a teacher when leading reflective discussions about students' sessions. We have built upon previous time series analysis tools to formulate a model that automatically segments a time series into these salient periods. The efficacy of the algorithm was demonstrated on a synthetic dataset where it outperformed previous work at the task of assigning data to regimes. We used the algorithm's output to generate a short text description of the dynamics in an inferred period. We find that evaluators are independently able to validate the inferred changes between the automatically generated periods. This preliminary study demonstrates that it is possible to simplify a time series log into periods of activity that are human interpretable.

Our focus now rests on designing assistive tools for teachers that can facilitate their understanding of students' interactions in multi-participant immersive simulations. Moreover, our results suggest that the model should be capable of adapting the number of inferred regimes to the complexity of a given session. Fox et al. [6] explore a Bayesian non-parametric model which allows the data to dictate the number of regimes that are inferred. The application of this model to the CW data presents an attractive tool for remaining agnostic about the number of regimes that are present in a session. Another avenue for future research involves exploring the trade-off that is made between the predictive power of a model and the explanatory coherence that the model achieves. Wu et al. [24] have suggested a method for regularizing deep learning models to facilitate people's understanding of their predictions. This is an important balance to consider and one that we intend to consider in educational settings.

7. ACKNOWLEDGMENTS

This paper is based on work supported by the National Science Foundation under grant IIS-1623124. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Thank you to Prof. Leilah Lyons (NYSci; University of Illinois-Chicago) and Aditi Mallavarapu (University of Illinois-Chicago) for their input on this work; and NYSci for assisting with the data collection.

[5]Visualization available at `https://s3.amazonaws.com/essil-validation/index.html`.

8. REFERENCES

[1] C. Brady, M. Horn, U. Wilensky, A. Wagh, A. Hjorth, and A. Banerjee. Getting your drift–activity designs for grappling with evolution. In *Proceedings of International Conference of the Learning Sciences, ICLS*, volume 3, pages 1603–1604. International Society of the Learning Sciences, 2014.

[2] O. Cappé, E. Moulines, and T. Rydén. Inference in hidden markov models. In *Proceedings of EUSFLAT Conference*, pages 14–16, 2009.

[3] V. Colella. Participatory simulations: Building collaborative understanding through immersive dynamic modeling. *The journal of the Learning Sciences*, 9(4):471–500, 2000.

[4] C. Dede, T. A. Grotzer, A. Kamarainen, and S. J. Metcalf. Virtual reality as an immersive medium for authentic simulations. In *Virtual, Augmented, and Mixed Realities in Education*, pages 133–156. Springer, 2017.

[5] P. Esling and C. Agon. Time-series data mining. *ACM Computing Surveys (CSUR)*, 45(1):12, 2012.

[6] E. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky. Nonparametric bayesian learning of switching linear dynamical systems. In *Advances in Neural Information Processing Systems*, pages 457–464, 2009.

[7] Z. Ghahramani and G. E. Hinton. Variational learning for switching state-space models. *Neural computation*, 12(4):831–864, 2000.

[8] P. Giordani, R. Kohn, and D. van Dijk. A unified approach to nonlinearity, structural change, and outliers. *Journal of Econometrics*, 137(1):112–133, 2007.

[9] A. Gnoli, A. Perritano, P. Guerra, B. Lopez, J. Brown, and T. Moher. Back to the future: embodied classroom simulations of animal foraging. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction*, pages 275–282. ACM, 2014.

[10] B. Horst and K. Abraham. *Data mining in time series databases*, volume 57. World scientific, 2004.

[11] N. Ikoma, T. Higuchi, and H. Maeda. Tracking of maneuvering target by using switching structure and heavy-tailed distribution with particle filter method. In *Control Applications, 2002. Proceedings of the 2002 International Conference on*, volume 2, pages 1282–1287. IEEE, 2002.

[12] A. Ioannidou, A. Repenning, D. Webb, D. Keyser, L. Luhn, and C. Daetwyler. Mr. vetro: A collective simulation for teaching health science. *International Journal of Computer-Supported Collaborative Learning*, 5(2):141–166, 2010.

[13] A. Jasra, C. C. Holmes, and D. A. Stephens. Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science*, pages 50–67, 2005.

[14] C.-J. Kim, C. R. Nelson, et al. State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books*, 1, 1999.

[15] V. R. Lee and J. Drake. Quantified recess: design of an activity for elementary students involving analyses of their own movement data. In *Proceedings of the 12th international conference on interaction design and children*, pages 273–276. ACM, 2013.

[16] V. R. Lee and J. M. Thomas. Integrating physical activity data technologies into elementary school classrooms. *Educational Technology Research and Development*, 59(6):865–884, 2011.

[17] T. Moher, B. Uphoff, D. Bhatt, B. López Silva, and P. Malcolm. Wallcology: Designing interaction affordances for learner engagement in authentic science inquiry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 163–172. ACM, 2008.

[18] K. P. Murphy and S. Russell. Dynamic bayesian networks: representation, inference and learning. 2002.

[19] P. Patel, E. Keogh, J. Lin, and S. Lonardi. Mining motifs in massive time series databases. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pages 370–377. IEEE, 2002.

[20] R. H. Shumway and D. S. Stoffer. Time series analysis and its applications. *Studies In Informatics And Control*, 9(4):375–376, 2000.

[21] O. Smørdal, J. Slotta, T. Moher, M. Lui, and A. Jornet. Hybrid spaces for science learning: New demands and opportunities for research. In *International Conference of the Learning Sciences. Sydney, Australia.*

[22] N. Whiteley, C. Andrieu, and A. Doucet. Efficient bayesian inference for switching state-space models using discrete particle markov chain monte carlo methods. *arXiv preprint arXiv:1011.2437*, 2010.

[23] U. Wilensky and M. Resnick. Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and technology*, 8(1):3–19, 1999.

[24] M. Wu, M. C. Hughes, S. Parbhoo, M. Zazzi, V. Roth, and F. Doshi-Velez. Beyond sparsity: Tree regularization of deep models for interpretability. *arXiv preprint arXiv:1711.06178*, 2017.